

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

- (i) APPLICANT: Black Jr., Charles A.
- (ii) TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR ACTIVATING GENES OF INTEREST
- (iii) NUMBER OF SEQUENCES: 16
- (iv) CORRESPONDENCE ADDRESS:
  - (A) ADDRESSEE: W. Murray Spruill
  - (B) STREET: 3605 Glenwood Ave. Suite 310
  - (C) CITY: Raleigh
  - (D) STATE: NC
  - (E) COUNTRY: US
  - (F) ZIP: 27622
- (v) COMPUTER READABLE FORM:
  - (A) MEDIUM TYPE: Floppy disk
  - (B) COMPUTER: IBM PC compatible
  - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
  - (D) SOFTWARE: PatentIn Release #1.0, Version #1.30
- (vi) CURRENT APPLICATION DATA:
  - (A) APPLICATION NUMBER:
  - (B) FILING DATE:
  - (C) CLASSIFICATION:
- (viii) ATTORNEY/AGENT INFORMATION:
  - (A) NAME: Spruill, W. Murray
  - (B) REGISTRATION NUMBER: 32,943
  - (C) REFERENCE/DOCKET NUMBER: 5722-2
- (ix) TELECOMMUNICATION INFORMATION:
  - (A) TELEPHONE: 919 420 2202
  - (B) TELEFAX: 919 881 3175

## (2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 4279 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: other nucleic acid
  - (A) DESCRIPTION: /desc = "Recombinant molecule (Multiple Cloning Site/Kozack sequence/LacZ gene)"
- (ix) FEATURE:
  - (A) NAME/KEY: misc\_feature
  - (B) LOCATION: 1..64
  - (D) OTHER INFORMATION: /product= "Multiple Cloning Site"
- (ix) FEATURE:
  - (A) NAME/KEY: misc\_feature

(B) LOCATION: 65..79  
(D) OTHER INFORMATION: /function= "Consensus sequence of translation initiation"  
/product= "Kozack sequence"

## (ix) FEATURE:

(A) NAME/KEY: prim\_transcript  
(B) LOCATION: 80..4279  
(D) OTHER INFORMATION: /gene= "LacZ"  
/standard\_name= "Beta galactosidase"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

TTAATACGAC TCACTATAGG CTAGCCTCGA GAATTCACGC GTGGTACCTC TAGAGTCGAC 60  
CCGGGCCGCC GCCACCATGG CGCAGCACCA TGGCCTGAAA TAACCTCTGA AAGAGGAACT 120  
TGGTTAGGTA CCTTCTGAGG CGGAAAGAAC CAGCTGTGGA ATGTGTGTCA GTTAGGGTGT 180  
GGAAAGTCCC CAGGCTCCCC AGCAGGCAGA AGTATGCAAA GCATGCATCT CAATTAGTCA 240  
GCAACCAGGT GTGGAAAGTC CCAGGCTCC CCAGCAGGCA GAAGTATGCA AAGCATGCAT 300  
CTCAATTAGT CAGCAACCAT AGTCCCGCCC CTAACTCCGC CCATCCCGCC CCTAACTCCG 360  
CCCAGTTCCG CCCATTCTCC GCCCATGGC TGAATAATTT TTTTATTTTA TGCAGAGGCC 420  
GAGGCCGCT CGGCCTCTGA GCTATTCCAG AAGTAGTGAG GAGGCTTTTT TGGAGGCCTA 480  
GGCTTTTGCA AAAAGCTTGG GATCTCTATA ATCTCGCGCA ACCTATTTTC CCCTCGAACA 540  
CTTTTAAAGC CGTAGATAAA CAGGCTGGGA CACTTCACAT GAGCGAAAAA TACATCGTCA 600  
CCTGGGACAT GTTGACAGATC CATGCACGTA AACTCGCAAG CCGACTGATG CCTTCTGAAC 660  
AATGGAAAGG CATTATTGCC GTAAGCCGTG GCGTCTGGT ACCGGTGGGT GAAGACCAGA 720  
AACAGCACCT CGAACTGAGC CGCGATATTG CCCAGCGTTT CAACGCGCTG TATGGCGAGA 780  
TCGATCCCGT CGTTTTACAA CGTCGTGACT GGGAAAACCC TGGCGTTACC CAACTTAATC 840  
GCCTTGCAGC ACATCCCCCT TTCGCCAGCT GCGTAATAG CGAAGAGGCC CGCACCGATC 900  
GCCCTTCCCA ACAGTTGCGC AGCCTGAATG GCGAATGGCG CTTTGCCTGG TTTCCGGCAC 960  
CAGAAGCGGT GCCGGAAAGC TGGCTGGAGT GCGATCTTCC TGAGGCCGAT ACTGTCGTCG 1020  
TCCCCTCAAA CTGGCAGATG CACGGTTACG ATGCGCCCAT CTACACCAAC GTAACCTATC 1080  
CCATTACGGT CAATCCGCCG TTTGTTCCCA CGGAGAATCC GACGGGTGT TACTCGCTCA 1140  
CATTTAATGT TGATGAAAGC TGGCTACAGG AAGGCAGAC GCGAATTATT TTTGATGGCG 1200  
TTAACTCGGC GTTTCATCTG TGGTGCAACG GCGCTGGGT CGGTACGGC CAGGACAGTC 1260  
GTTTGCCGTC TGAATTTGAC CTGAGCGCAT TTTTACGCGC CGGAGAAAAC CGCCTCGCGG 1320  
TGATGGTGCT GCGTTGGAGT GACGGCAGTT ATCTGGAAGA TCAGGATATG TGGCGGATGA 1380  
GCGGCATTTT CCGTGACGTC TCGTTGCTGC ATAAACCGAC TACACAAATC AGCGATTTCC 1440

ATGTTGCCAC TCGCTTTAAT GATGATTTCA GCCGCGCTGT ACTGGAGGCT GAAGTTCAGA 1500  
TGTGCGGCGA GTTGCCTGAC TACCTACGGG TAACAGTTTC TTTATGGCAG GGTGAAACGC 1560  
AGGTCGCCAG CGGCACCGCG CTTTTCGGCG GTGAAATTAT CGATGAGCGT GGTGGTTATG 1620  
CCGATCGCGT CACACTACGT CTGAACGTCG AAAACCCGAA ACTGTGGAGC GCCGAAATCC 1680  
CGAATCTCTA TCGTGCGGTG GTTGAAGTGC ACACCGCCGA CGGCACGCTG ATTGAAGCAG 1740  
AAGCCTGCGA TGTGCGTTTC CGCGAGGTGC GGATTGAAAA TGGTCTGCTG CTGCTGAACG 1800  
GCAAGCCGTT GCTGATTCGA GCGCTTAACC GTCACGAGCA TCATCCTCTG CATGGTCAGG 1860  
TCATGGATGA GCAGACGATG GTGCAGGATA TCCTGCTGAT GAAGCAGAAC AACTTTAACG 1920  
CCGTGCGCTG TTCGCATTAT CCGAACCATC CGCTGTGGTA CACGCTGTGC GACCGCTACG 1980  
GCCTGTATGT GGTGGATGAA GCCAATATTG AAACCCACGG CATGGTGCCA ATGAATCGTC 2040  
TGACCGATGA TCCGCGCTGG CTACCGGCGA TGAGCGAACG CGTAACGCGA ATGGTGCAGC 2100  
GCGATCGTAA TCACCCGAGT GTGATCATCT GGTGCTGGG GAATGAATCA GGCCACGGCG 2160  
CTAATCACGA CGCGCTGTAT CGCTGGATCA AATCTGTCTG TCCTTCCCGC CCGGTGCAGT 2220  
ATGAAGGCGG CGGAGCCGAC ACCACGGCCA CCGATATTAT TTGCCCGATG TACGCGCGCG 2280  
TGGATGAAGA CCAGCCCTTC CCGGCTGTGC CGAAATGGTC CATCAAAAAA TGGCTTTTCG 2340  
TACCTGGAGA GACGCGCCCG CTGATCCTTT GCGAATACGC CCACGCGATG GGTAACAGTC 2400  
TTGGCGGTTT CGCTAAATAC TGGCAGGCGT TTCGTCAGTA TCCCCGTTTA CAGGGCGGCT 2460  
TCGTCTGGGA CTGGGTGGAT CAGTCGCTGA TTAAATATGA TGAAAACGGC AACCCGTGGT 2520  
CGGCTTACGG CGGTGATTTT GGCATACGC CGAACGATCG CCAGTTCTGT ATGAACGGTC 2580  
TGGTCTTTGC CGACCGCACG CCGCATCCAG CGCTGACGGA AGCAAAACAC CAGCAGCAGT 2640  
TTTTCCAGTT CCGTTTATCC GGGCAAACCA TCGAAGTGAC CAGCGAATAC CTGTTCCGTC 2700  
ATAGCGATAA CGAGCTCCTG CACTGGATGG TGGCGCTGGA TGTAAGCCG CTGGCAAGCG 2760  
GTGAAGTGCC TCTGGATGTC GCTCCACAAG GTAAACAGTT GATTGAACTG CCTGAACTAC 2820  
CGCAGCCGGA GAGCGCCGGG CAACTCTGGC TCACAGTACG CGTAGTGCAA CCGAACGCGA 2880  
CCGCATGGTC AGAAGCCGGG CACATCAGCG CCTGGCAGCA GTGGCGTCTG GCGGAAAACC 2940  
TCAGTGTGAC GCTCCCCGCC GCGTCCCACG CCATCCCGCA TCTGACCACC AGCGAAATGG 3000  
ATTTTTCAT CGAGCTGGGT AATAAGCGTT GGCAATTTAA CCGCCAGTCA GGCTTTCTTT 3060  
CACAGATGTG GATTGGCGAT AAAAAACAAC TGCTGACGCC GCTGCGCGAT CAGTTCACCC 3120  
GTGCACCGCT GGATAACGAC ATTGGCGTAA GTGAAGCGAC CCGCATTGAC CCTAACGCCT 3180  
GGGTCGAACG CTGGAAGGCG GCGGGCCATT ACCAGGCCGA AGCAGCGTTG TTGCAGTGCA 3240

CGGCAGATAC ACTTGCTGAT GCGGTGCTGA TTACGACCGC TCACGCGTGG CAGCATCAGG 3300  
GGAAAACCTT ATTTATCAGC CGGAAAACCT ACCGGATTGA TGGTAGTGGT CAAATGGCGA 3360  
TTACCGTTGA TGTGGAAGTG GCGAGCGATA CACCGCATCC GCGCGGATT GGCCTGAACT 3420  
GCCAGCTGGC GCAGGTAGCA GAGCGGGTAA ACTGGCTCGG ATTAGGGCCG CAAGAAAACCT 3480  
ATCCCGACCG CCTTACTGCC GCCTGTTTTG ACCGCTGGGA TCTGCCATTG TCAGACATGT 3540  
ATACCCCGTA CGTCTTCCCG AGCGAAAACG GTCTGCGCTG CGGGACGCGC GAATTGAATT 3600  
ATGGCCCACA CCAGTGGCGC GGCGACTTCC AGTTCAACAT CAGCCGCTAC AGTCAACAGC 3660  
AACTGATGGA AACCAGCCAT CGCCATCTGC TGCACGCGGA AGAAGGCACA TGGCTGAATA 3720  
TCGACGGTTT CCATATGGGG ATTGGTGGCG ACGACTCCTG GAGCCCGTCA GTATCGGCGG 3780  
AATTCCAGCT GAGCGCCGGT CGCTACCATT ACCAGTTGGT CTGGTGTCAA AAATAATAAT 3840  
AACCGGGCAG GCCATGTCTG CCCGTATTTT GCGTAAGGAA ATCCATTATG TACTATTTAA 3900  
AAAACACAAA CTTTGGATG TTCGGTTTAT TCTTTTTCTT TTACTTTTTT ATCATGGGAG 3960  
CCTACTTCCC GTTTTCCCG ATTTGGCTAC ATGACATCAA CCATATCAGC AAAAGTGATA 4020  
CGGGTATTAT TTTTGCCGCT ATTTCTCTGT TCTCGCTATT ATTCCAACCG CTGTTTGGTC 4080  
TGCTTTCTGA CAAACTCGGA ACTTGTTTAT TGCAGCTTAT AATGGTTACA AATAAAGCAA 4140  
TAGCATCACA AATTCACAA ATAAAGCATT TTTTCACTG CATTCTAGTT GTGGTTTGTG 4200  
CAAATCATC AATGTATCTT ATCATGTCTG GATCCTCTAG AGTCGACCTG CAGGCATGCA 4260  
AGCTGGCACT GGCCGTCGT 4279

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "Synthetic oligonucleotide"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

GAATACAAAG CTTATGCATG

20

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 13 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: other nucleic acid  
(A) DESCRIPTION: /desc = "Synthetic oligonucleotide"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

GAATACAAAG CTT

13

(2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 20 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: other nucleic acid  
(A) DESCRIPTION: /desc = "Synthetic oligonucleotide"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

AAAGCTTATG CATGCGCCG

20

(2) INFORMATION FOR SEQ ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 20 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: other nucleic acid  
(A) DESCRIPTION: /desc = "Synthetic oligonucleotide"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

CGGCCGCATC TAGAGGGCCC

20

(2) INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 25 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: other nucleic acid  
(A) DESCRIPTION: /desc = "Synthetic oligonucleotide"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

GCGGCCGCAT CTAGAGGGCC CGGAT

25

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "Synthetic oligonucleotide"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

AATACAAAGC TTATGCATGC GGCC

24

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "Synthetic oligonucleotide"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

AATACAAAGC TTATGCATGC GGCCGCATCT

30

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "Synthetic oligonucleotide"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

CATGCATAAG CTTTGTATTC

20

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 13 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "Synthetic oligonucleotide"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

AAGCTTTGTA TTC

13

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "Synthetic oligonucleotide"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

CGGCCGCATG CATAAGCTTT

20

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "Synthetic oligonucleotide"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

GGGCCCTCTA GATGCGCCG

20

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "Synthetic oligonucleotide"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

ATCCGGGCCC TCTAGATGCG GCCGC

25

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "Synthetic oligonucleotide"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

GGCCGCATGC ATAAGCTTTG TATT

24

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "Synthetic oligonucleotide"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

AGATGCGGCC GCATGCATAA GCTTTGTATT

30

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1798 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: mRNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

GAAUACAAAG CUUAUGCAUG CGGCCGCAUC UAGAGGGCCC GGAUCCAAAU GGAAGACGCC

60



AAAAACAUAU AGAAAGGCCG GCGCCAUUC UAUCCUCUAG AGGAUGGAAC CGCUGGAGAG 120  
CAACUGCAUA AGGCUAUGAA GAGAUACGCC CUGGUUCCUG GAACAAUUGC UUUUACAGAU 180  
GCACAUUUCG AGGUGAACAU CACGUACGCG GAAUACUUCG AAAUGUCCGU UCGGUUGGCA 240  
GAAGCUAUGA AACGAUAUGG GCUGAAUACA AAUCACAGAA UCGUCGUAUG CAGUGAAAAC 300  
UCUCUUCAAU UCUUUAUGCC GGUGUUGGGC GCCGUUAUUU AUCGGAGUUG CAGUUGCGCC 360  
CGCGAAGCAC AUUUUAUAAUG AACGUGAAUU GCUCAACAGU AUGAACAUUU CGCAGCCUAC 420  
CGUAGUGUUU GUUUCCAAAA AGGGGUUGCA AAAAAUUUUG AACGUGCAAA AAAAAUACC 480  
AAUAAUCCAG AAAAUUAUUA UCAUGGAUUC UAAAACGGAU UACCAGGGAU UUCAGUCGAU 540  
GUACACGUUC GUCACAUCUC AUUACCUCC CGGUUUUAU GAAUACGAU UUGUACCAGA 600  
GUCCUUUGAU CGUGACAAAA CAUUGCACU GAUAAUGAAU UCCUCUGGAU CUACUGGGUU 660  
ACCUAAGGGU GUGGCCCUUC CGCAUAGAAC UGCCUGCGUC AGAUUCUCGC AUGCCAGAGA 720  
UCCUAUUUUU GGCAAUCAA UCAUUCGGGA UACUGCGAUU UUAAGUGUUG UUCAUUGCA 780  
UCACGGUUUU GGAAUGUUUA CUACACUCCG AUUUUUGAU UGUGGAUUUC GAGUCGUCUU 840  
AAUGUAUAGA UUGAAGAAG AGCUGUUUUU ACGAUCCCUU CAGGAUUACA AAAUUCAAAG 900  
UGCGUUGCUA GUACCAACCC UAUUUUCAU CUUCGCCAAA AGCACUCUGA UUGACAAAUA 960  
CGAUUUUUCU AAUUUACACG AAAUUGCUUC UGGGGCGCA CCUCUUUCGA AAGAAGUCGG 1020  
GGAAGCGGUU GCAAACGCU UCCAUCUUC AGGGAUACGA CAAGGAUAUG GGCUCACUGA 1080  
GACUACAUA GCUAUUCUGA UUACACCCGA GGGGGAUGAU AAACCGGGCG CGGUCGGUAA 1140  
AGUUGUCCA UUUUUUGAAG CGAAGGUUGU GGAUCUGGAU ACCGGGAAA CGCUGGGCGU 1200  
UAAUCAGAGA GGCGAUUUAU GUGUCAGAGG ACCUAUGAUU AUGUCCGGUU AUGUAAACAA 1260  
UCCGGAAGCG ACCAACGCCU UGAUUGACAA GGAUGGAUGG CUACAUUCUG GAGACAUAGC 1320  
UUACUGGGAC GAAGACGAAC ACUUCUUAU AGUUGACCGC UUGAAGUCUU UAAUUAAAUA 1380  
CAAAGGAUUA CAGGUGGCC CCGUGAAUU GGAAUCGAUA UUGUUACAAC ACCCAACAU 1440  
CUUCGACGCG GCGUGGCAG GUCUCCCGA CGAUGACGCC GGUGAACUUC CCGCCGCCGU 1500  
UGUUGUUUG GAGCACGGAA AGACGAUGAC GGAAAAGAG AUCGUGGAU ACGUCGCCAG 1560  
UCAAGUAACA ACCGCGAAAA AGUUGCGCG AGGAGUUGUG UUGUGGACG AAGUACCGAA 1620  
AGGUCUUACC GGAAACUCG ACGCAAGAAA AAUCAGAGAG AUCCUCAUA AGGCCAAGAA 1680  
GGGCGGAAAG UCCAAAUUGU AAAAUGUAA UGUUUCAGC GAUGACGAAA UUCUUAGCUA 1740  
UUGUAAUCCU CCGAGGGGGC GAGCUCCCAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 1798